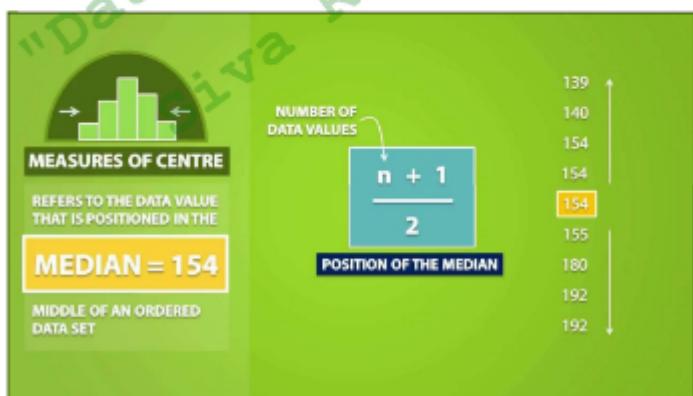
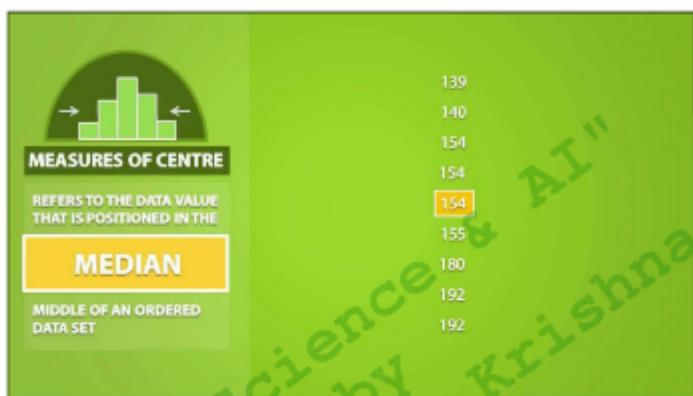


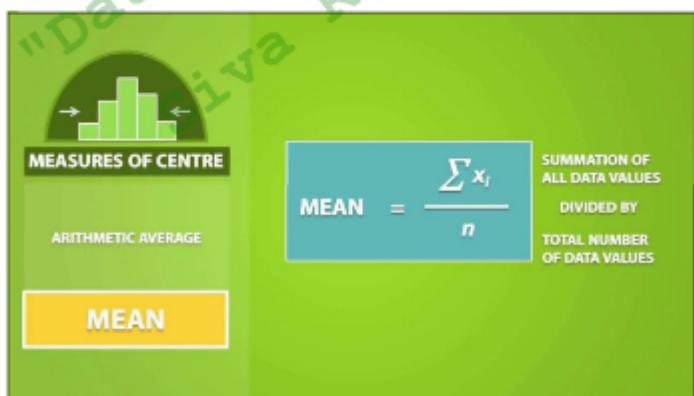


MEASURES of CENTRAL TENDENCY











Population Mean

$$\begin{aligned}\mu &= \frac{\sum X}{N} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{N} \\ &= \frac{24 + 13 + 19 + 26 + 11}{5} \\ &= \frac{93}{5} \\ &= 18.6\end{aligned}$$

Sample Mean

$$\begin{aligned}\bar{X} &= \frac{\sum X}{n} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \\ &= \frac{57 + 86 + 42 + 38 + 90 + 66}{6} \\ &= \frac{379}{6} \\ &= 63.167\end{aligned}$$



Weighted Average

- Sometimes we wish to average numbers, but we want to assign more importance, or weight, to some of the numbers.
- The average you need is the weighted average.

$$\text{Weighted Average} = \frac{\sum xw}{\sum w}$$

where x is a data value and w is the weight assigned to that data value. The sum is taken over all data values.

Example

Suppose your midterm test score is 83 and your final exam score is 95. Using weights of 40% for the midterm and 60% for the final exam, compute the weighted average of your scores. If the minimum average for an A is 90, will you earn an A?

$$\begin{aligned}\text{Weighted Average} &= \frac{(83)(0.40) + (95)(0.60)}{0.40 + 0.60} \\ &= \frac{32 + 57}{1} = 90.2\end{aligned}$$

You will earn an A!

Various Sampling Techniques

- **Random sampling**

- Every unit of the population has the same chance of being included in the sample.
- A chance mechanism is used in the selection process.
- Eliminates bias in the selection process.
- Also known as probability sampling

- **Nonrandom Sampling**

- Every unit of the population does not have an equal chance of being included in the sample.
- Open to the selection bias
- Not appropriate data collection method
- Also known as non-probability sampling

Simple Random Sampling

- Every object in the population has an equal chance of being selected
- Objects are selected independently of each other
- Samples can be obtained from a table of random numbers or by using random number generators
- A simple random sample is the ideal type of sample because all objects in the population are equally likely to be included and all samples of the same size have an equal chance of being selected

Systematic Sampling

- Convenient and relatively easy to administer
- Population elements are an ordered sequence (at least, conceptually).
- The first sample element is selected randomly from the first k population elements.
- Thereafter, sample elements are selected at a constant interval, k , from the ordered sequence frame.

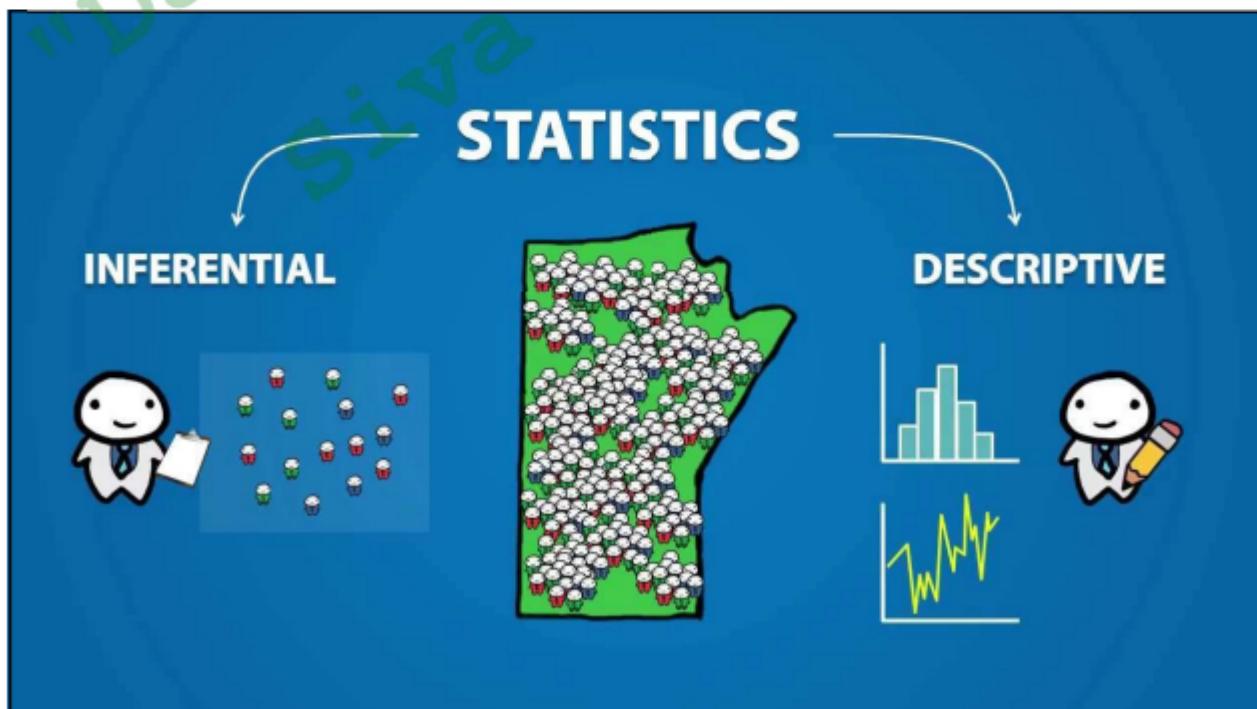
Nonrandom Sampling

- **Convenience Sampling:** Sample elements are chosen based on convenience of the researcher
- **Judgment Sampling:** Sample elements are chosen based on the judgment of the researcher
- **Quota Sampling:** Sample elements are chosen until specific quotas are satisfied
- **Snowball Sampling:** Survey subjects are recruited from other survey respondents



STATISTICS

MEASURE + ANALYZE





Descriptive vs Inferential Statistics

- **Descriptive statistics**
 - Collecting, presenting, and describing data
- **Inferential statistics**
 - Drawing conclusions and/or making decisions concerning a population based only on sample data



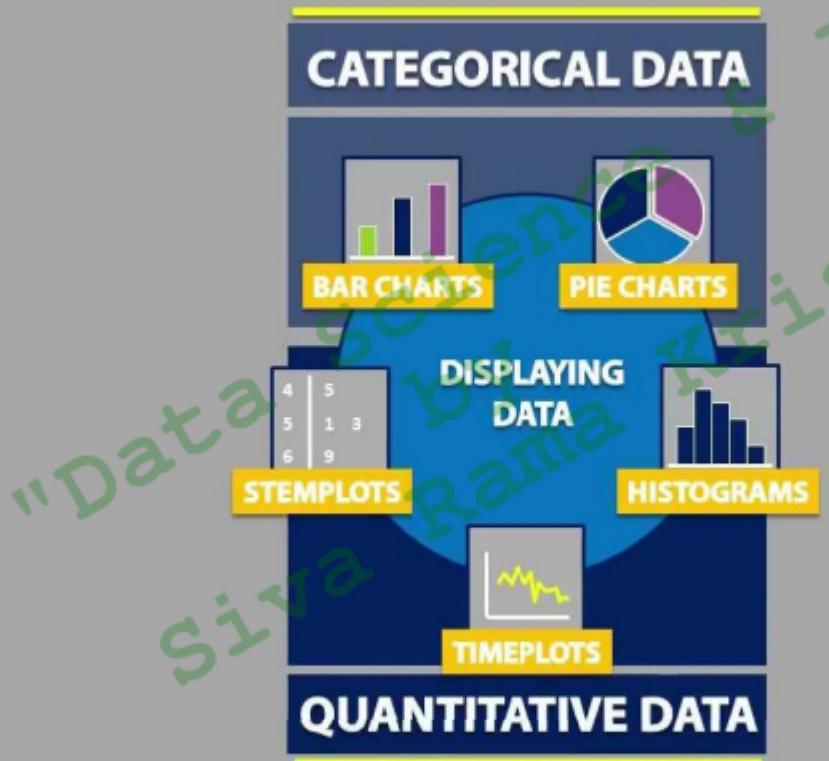
Populations and Samples

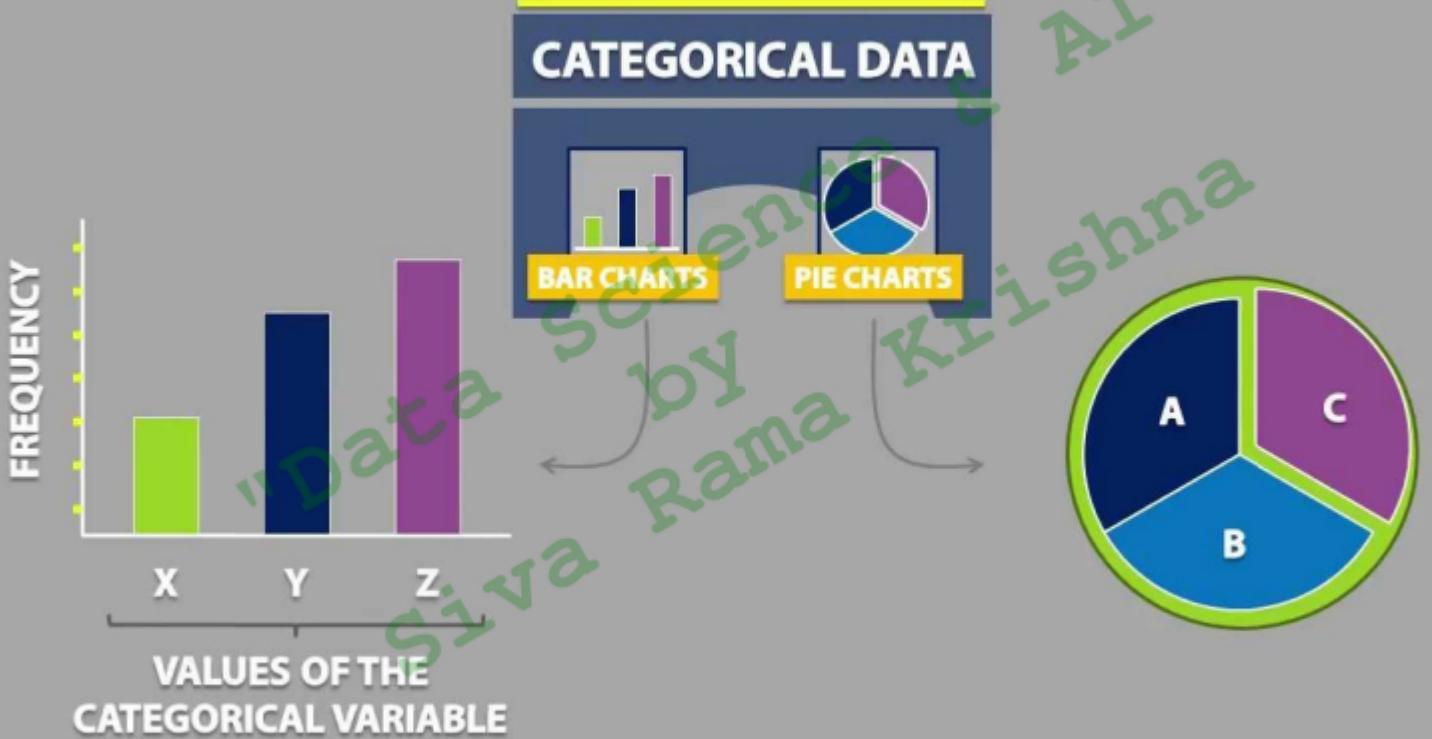
- A **Population** is the set of all items or individuals of interest
 - Examples: All likely voters in the next election
All parts produced today
All sales receipts for November
- A **Sample** is a subset of the population
 - Examples: 1000 voters selected at random for interview
A few parts selected for destructive testing
Random receipts selected for audit

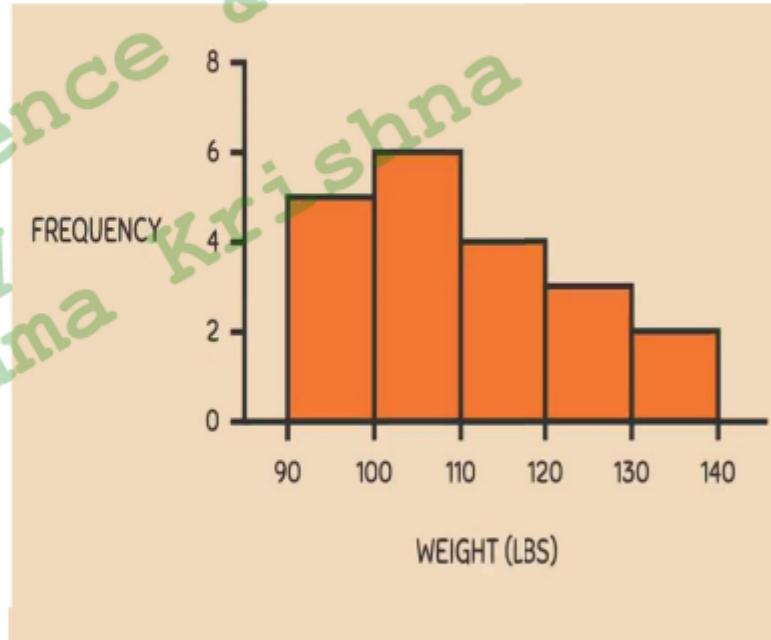
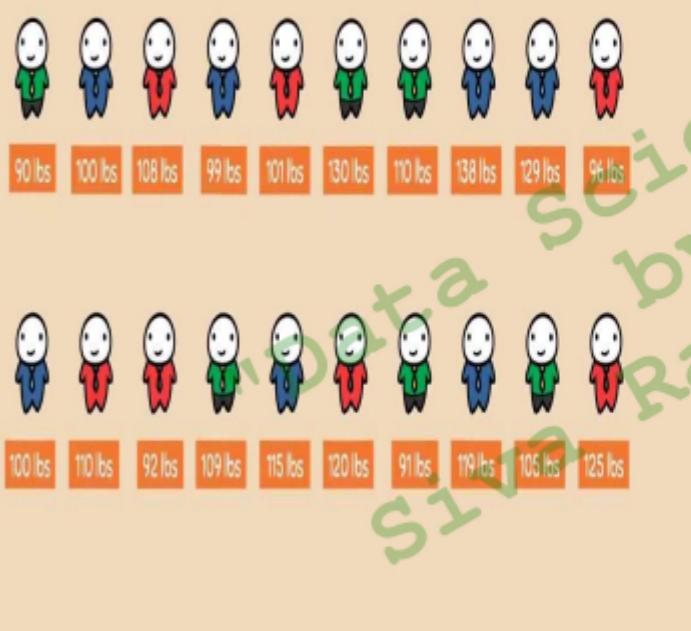


Why Sample?

- Less time consuming than a census
- Less costly to administer than a census
- It is possible to obtain statistical results of a sufficiently high precision based on samples.
- Because the research process is sometimes destructive, the sample can save product
- If accessing the population is impossible; sampling is the only option









FREQUENCY DISTRIBUTION



WEIGHT

100 – 110	
110 – 120	120
120 – 130	130
130 – 140	
140 – 150	

BY CONVENTION, WE SAY THAT
EACH INTERVAL DOES NOT
INCLUDE THE RIGHT END POINT

-
-
-



"Data Science & AI"

Siva Ranjan Krishna

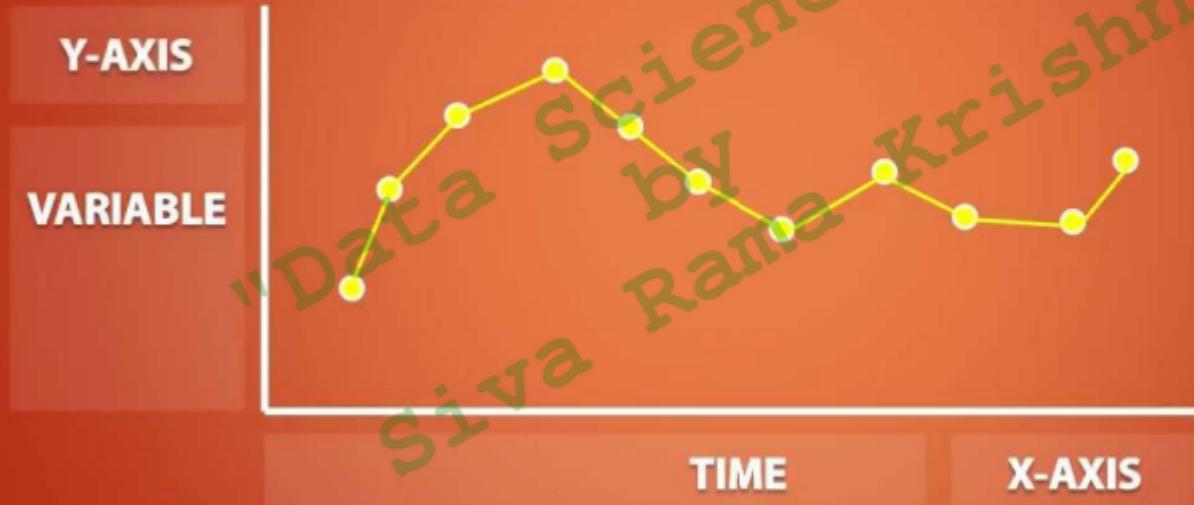
TIMEPLOT

SHOW HOW A VARIABLE CHANGES OVER TIME



TIMEPLOT

SHOW HOW A VARIABLE CHANGES OVER TIME





VARIABILITY

FOOD PREFERENCES



"Data Science & AI" by Ramz, Siva, Krishnan

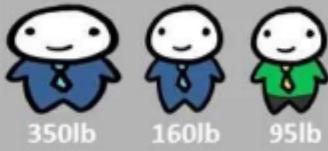
HAIR COLOUR



HEIGHT



WEIGHT





MEASURE

VARIABLE

QUANTITATIVE DATA

DATA THAT IS MEASURED IN NUMBERS. IT DEALS WITH NUMBERS THAT MAKE SENSE TO PERFORM ARITHMETIC CALCULATIONS WITH

QUANTITATIVE VARIABLES

HEIGHT

WEIGHT

MIDTERM SCORE

CATEGORICAL DATA

REFERS TO THE VALUES THAT PLACE "THINGS" INTO DIFFERENT GROUPS OR CATEGORIES

CATEGORICAL VARIABLES

HAIR COLOUR

TYPE OF CAT

LETTER GRADE



CATEGORICAL VARIABLE

CATEGORICAL AND ORDINAL

LOGICAL ORDERING TO THE VALUES OF A CATEGORICAL VARIABLE

EX: LETTER GRADE

F C C+ B B+ A A+

CATEGORICAL AND NOMINAL

NO LOGICAL ORDERING TO THE VALUES OF A CATEGORICAL VARIABLE

EX: HAIR COLOUR

RED BLONDE BROWN BLUE



QUANTITATIVE VARIABLE

DISCRETE

REFER TO VARIABLES THAT CAN ONLY
BE MEASURED IN CERTAIN NUMBERS

CONTINUOUS

REFER TO VARIABLES THAT CAN TAKE
ON ANY NUMERICAL VALUE

EX: NUMBER OF PETS YOU OWN

0 1 2 30 2.7

EX: WEIGHT

105 185 170.683

